Some Applications of the Difference Analysis for Stochastic Systems

A. Yu. Shahverdian (1) and A. V. Apkarian (2)

(1) Yerevan Physics Institute, Yerevan, Armenia
(2) Northwestern University, The Medical School, Chicago, IL, U.S.A.

Abstract

The work relates to a new way for analysis of one-dimensional stochastic systems, based on consideration of its higher order difference structure. From this point of view, the deterministic and random processes are analyzed. A new numerical characteristic for one-dimensional stochastic systems is introduced. The applications to single neuron models and neural networks are given.

1 Introduction

This paper presents some applications of the difference analysis, has been suggested in authors' recent works [1-3]. The approach has been detected on computational study [1] of neural activity: we observed that sequences of higher order absolute differences, taken from periodically stimulated neuron's spike train, contain long samples on which the changes in monotony (increase/decrease) are periodic.

The next Section 2 generalizes this observation and introduces a new characteristic for stochastic one-dimensional systems, expressing this type of non-explicit periodicity in a quantitative form. This notion, the numerical characteristic γ, being a measure of some minimal periodicity, can also be treated as a new measure of irregularity. On instances of some nonlinear maps its numerical comparisons with Lyapunov exponent are given.

The difference approach provides a strict algorithmic base for detection of small chaotic fluctuations at unstable bifurcation points. This is given for logistic map, however the same analysis can be applied for various systems, exhibiting the bifurcations.

The probabilistic systems permit rigorous study – Section 3 considers discrete random processes. The sequences of independent random variables as well as binary Markov processes are examined. In fact, we deal with a new type of limiting transition, taken over the discrete random process. The applications to diffusion and Poisson processes are given. We establish several theoretical results relating to γ-characteristic of random discrete systems. We are based on Eggleston theorem [7] from ergodic theory to describe the attractors for these random processes.

Section 4 considers stochastic models of single neuron as well as establishes some probabilistic neuron's firing condition. We are again based on the difference analysis and use a modified version of entropy. We claim that some new type of attractors, resulting in the approach suggested, can be treated as a new extended memory in neural networks. This appears to be well consistent with the brain theory concepts of attractor computing and associative hierarchical memory [8]. Finally, we prove a theorem, showing that if follow the difference approach, the noise in fact can be eliminated from arbitrary "noisy" neural network.

The paper is organized in such a way, that the most theoretical results can be easily deduced from the previous ones. The complete proofs of the other theoretical statements will appear in the version of this work submitted to publication to a regular journal.

2 Deterministic processes

2.1 Finite differences and conjugate orbits

The difference approach, suggested in [1-3], reduces the study of stochastic properties of the orbits $\bar{X} = (x_i)_{i=1}^{\infty}$, generated by given one-dimensional system, to analysis of alternations of the monotone increase or decrease of higher order absolute differences

$$\Delta^{(s)}x_i = |\Delta^{(s-1)}x_{i+1} - \Delta^{(s-1)}x_i| \qquad (\Delta^{(0)}x_i = x_i\ ;\ \ i, s = 1, 2, 3, \ldots)\ .$$

In this section we describe some statements of the approach and explain some basic notions involved in this work.

Let us have a one-dimensional stochastic system, generating numerical sequences $\bar{X} = (x_i)_{i=1}^{\infty},\ 0 \leq x_i \leq 1$. It is not difficult to see, that for $1 \leq s \leq k-1$ we have

$$\Delta^{(s-1)}x_i = \mu_{k,s-1} + \sum_{p=1}^{i-1}(-1)^{\delta_p^{(s)}}\Delta^{(s)}x_p - \min_{0\leq i\leq k-s}(\sum_{p=1}^{i}(-1)^{\delta_p^{(s)}}\Delta^{(s)}x_p) \tag{1}$$

where

$$\delta_p^{(s)} = \begin{cases} 0 & \Delta^{(s)}x_{p+1} \geq \Delta^{(s)}x_p \\ 1 & \Delta^{(s)}x_{p+1} < \Delta^{(s)}x_p \end{cases} \qquad \mu_{k,s} = \min\{\Delta^{(s)}x_i : 1 \leq i \leq k-s\}\ ,$$

(it is assumed $\sum_1^0 = 0$). Using the recurrent formula (1) we transform the finite orbit $\bar{X}_k$ into some special form, which emphasizes its higher order difference structure. Namely, we transform $\bar{X}_k = (x_i)_{i=1}^{k}$ into the sequence $\bar{\zeta}_k$,

$$\bar{\zeta}_k = (\bar{\lambda}_k,\ \bar{\mu}_k,\ \bar{\rho}_k) \tag{2}$$

where

$$\bar{\lambda}_k = (\lambda_1, \lambda_2, \ldots, \lambda_k), \qquad \lambda_s = 0.\delta_1^{(s)}\delta_2^{(s)}\ldots, \delta_{k-s}^{(s)} \qquad (s = 0, 1, \ldots, m) \tag{3}$$

$$\bar{\mu}_k = (\mu_{k,1},\ \mu_{k,2},\ \ldots, \mu_{k,m}), \qquad \bar{\rho}_k = (\rho_{k,1},\ \rho_{k,2},\ \ldots, \rho_{k,k-m}) \quad (\rho_{k,i} = \Delta_i^{(m)})\ . \tag{4}$$

The Eq. (2) in fact represents the original orbit in a different form – one can see that applying the recurrent procedure (1), the sequence $\bar{X}_k$ can be completely recovered by $\bar{\zeta}_k$. The difference method suggests to study the sequences $\bar{\nu} = (\nu_k)_{k=1}^{\infty}$ which terms are defined as follows:

$$\nu_k = 0.\delta_1^{(k)}\delta_2^{(k)}\delta_3^{(k)}\ldots \tag{5}$$

– it is clear that

$$|\nu_k - \lambda_k| \leq 2^{-k} \qquad (k \geq 1)\ . \tag{6}$$

The $\bar{\nu}$ is called the conjugate (to $\bar{X}$) orbit. The approach distinguishes two cases – continuous, when the quantities $\rho_{k,i}$ from (4) can take arbitrary numerical values from interval $(0, 1)$, and the discrete case, when they take only a finite number of values. In either case, given $\bar{X}$ we are interested in its higher order difference structure, that reflects the conjugate orbit $\bar{\nu}$.

The computations performed in [1-2] show that for many actual continuous-time systems the quantities

$$||\bar{\mu}_k||^2 + ||\bar{\rho}_k||^2 \qquad (||(x_1, \ldots, x_m)|| = (\sum_{i=1}^{m} x_i^2)^{1/2}) \tag{7}$$

as well as (due to relations (2) and (6)) the distances $||\bar{\zeta}_k - \bar{\nu}_k||$ converge to zero with the exponential rate. In contrary, the sequences $\bar{\nu}_k$ from (5) mostly have oscillating character and are attracted either to an interval or to a thin set $\mathcal{A}$. This means, that the dominant part of information, that carries the original time series $\bar{X}$ and which permits measurements, in fact is conveyed by its conjugate orbit $\bar{\nu}$. For many irregular systems the set $\mathcal{A}$ is the same for different orbits determining by different initial states. Hence, the $\mathcal{A}$ appears to be some (conjugate) attractor for the system. Therefore, the main part of information, produced by system during its evolution, is contained in the attractor $\mathcal{A}$, which can be treated as a geometrical image of the whole produced information.

The systems, generating some natural numbers belonging to a finite set, (e.g., as the outcomes in hazard games) should be classified to the discrete case. The conjugate orbits are constructed by the same way: for instance, if $\bar{X} = (x_i)_{i=0}^{\infty}$ and $x_i \in \{0, 1\}$, then the difference sequences $\bar{X}^{(k)} = (x_i^{(k)})_{i=0}^{\infty}$ are again the binary sequences and the conjugate orbits consist of the terms

$$\nu_k = \sum_{n=1}^{\infty} 2^{-n} x_n^{(k)} \ .$$

However, for this case we do not have the convergence of (7) to zero as for continuous-time systems. Instead, prescribing some probabilities to generated outcomes (i.e. considering the random sequences) we are able to compare the original and conjugate systems just through their analytical parameters (see Section 3 for details). The sequences ν_k, being considered on some infinite subsets of indices $\Lambda \subset N$ of the natural series N, are convergent to some compacts from $(0, 1)$, which can be disjoint for different Λ. For the sequences of independent random variables and Markov chains these cluster sets permit analytical description. Moreover, the analysis suggested can be applied to arbitrary continuous-time stochastic systems, permitting approximation or interpolation by the discrete ones. E.g., Section 3 gives such an application to diffusion stochastic processes. As another example (does not considered in this work) one can refer to Nyquist-Shannon sampling theorem [8, 32] from data analysis, according to which the analogue signal with bounded power spectrum is completely determined through its values on some discrete set of equidistant points.

2.2 A new characteristic for irregular systems

One of the main claims relating to the approach described, which has been confirmed by preliminary computations ([3]), is the following: a periodic response of given stochastic

system on a weak periodic perturbation is localized in $\bar{X}^{(n)}$ – its presence can be detected, when considering the higher order differences, taken from initial orbit X. For that purpose it should be studied the asymptotical (as $N \to \infty$) relative volume (the density in natural series) of the set of all those indices i, for which the changes of binary symbol occur,

$$\delta_{i+1}^{(N)} = 1 - \delta_i^{(N)} \qquad (1 \leq i \leq N-1). \tag{8}$$

This leads to the following definition: For an orbit $\bar{X} = (x_k)_{k=0}^{\infty}$ of a given system we define

$$\gamma = \gamma(\bar{X}) = \lim_{N\to\infty} \frac{\gamma(\bar{X}, N)}{N} \qquad (0 \leq \gamma \leq 1) \tag{9}$$

where $\gamma(\bar{X}, N)$ denotes the total number of those indices $1 \leq i \leq N-1$ for each of which the (8) holds; the existence of limit in (9) is preassumed.

For deterministic systems the theoretical study of properties of γ is quite difficult. Two statements of this section, Theorem 1 and Corollary 1, are apparently the only available theoretical results. Some theoretical statements, relating to random binary sequences can be found in Section 3. On the other hand, the γ has that important advantage, that is the simplicity of its computation. It can be easily implemented just over the (experimental) data $\bar{X}$, without referring to the process generation law. By this reason, as for computation of γ we need only the corresponding time series to be available, this new characteristic is well adapted for computational study of various applied problems.

The numerical analysis shows (see [3]) that γ has weak dependence on system's initial values and is able to distinguish the regular and chaotic motion. With this aim we have compared γ with Lyapunov exponent λ (see e.g., [12], Ch. 5 and [4], Ch. 7.2.b): for a given map $F : (0,1) \to (0,1)$ it is defined as

$$\lambda = \lambda(\bar{X}) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \ln \left| \frac{dF(x_k)}{dx_k} \right| \qquad (x_{k+1} = F(x_k)) \ . \tag{10}$$

It is known [4] that Lyapunov exponent of any integrable system is zero. In contrary, it follows from Theorem 1 below that there exist integrable (but aperiodic) systems with positive (and rational) γ – it can be, e.g., the sequence of fractional parts. Some results on computations of γ-characteristic as well as its comparisons with Lyapunov exponent can be found in [3]. Three simple deterministic systems – the tent map, logistic function, and Poincare displacement of Chirikov's standard map have been examined. The numerical results demonstrate a strong correlation of γ with Lyapunov exponent (see [3], Figs. 1 and 2).

We emphasize another important feature of this quantity, has been derived from these preliminary computations: the γ-coefficient is that numerical characteristic, associated with a given stochastic system, which is able to change significantly its numerical value when the system undergoes a weak perturbation. This relates to stochastic resonance phenomena. Some works [24, 25, 26] claim that namely this resonance mechanism has the basic role in neural activity.

We have proven two rigorous results relating to γ-characteristic of continuous deterministic systems – Theorem 1 and Corollary 1. The Corollary 1 follows from Theorem

2, giving some 'difference' analogy of Eggleston formula from the ergodic theory. If $\bar{X}$ is either constant or periodic, then we obviously have

$$\gamma(\bar{X}, N) = \gamma N + O(1) \qquad (N \to \infty) \tag{11}$$

and the coefficient $0 \le \gamma \le 1$ is rational. According to next theorem (the symbols $\{.\}$ and $[.]$ denote fractional and entire part of number) this remains valid also for sequences of fractional parts. These (conditionally periodic) sequences have an important role on studying the general integrable dynamical systems (see [34, 35]).

Theorem 1 *For the sequence $\bar{X} = (\{\alpha n\})_{n=1}^{\infty}$ where $0 < \alpha < 1$ is irrational, the next statements are true: (1) the conjugate to $\bar{X}$ orbit $\bar{\nu} = (\nu_n)_{n=1}^{\infty}$ is a periodic sequence; (2) if entire part of $1/\alpha$ is of the form $[1/\alpha] = 2^p - 1$ $(p \ge 1)$ then $\nu_n \equiv 0$ for all large enough indices n.*

The second theoretical result on γ-characteristic is the Corollary 1, establishing for binary systems an upper estimate for Hausdorff dimension of the attractors $\mathcal{A}$ (see previous section 2.1). To formulate this estimate involving γ-characteristic and Shannon entropy function, we need some preliminary definitions and results. Let us consider the processes, generating the binary sequences

$$\bar{x} = (x_1, x_2, \ldots, x_n, \ldots), \qquad x_i \in \{0, 1\} \; ; \tag{12}$$

it is convenient to prescribe to such a sequence the number $0 < x < 1$

$$x = 0.x_1 x_2 \ldots x_n \ldots \qquad (= \sum_{n=1}^{\infty} 2^{-n} x_n) \; .$$

We define $\mathcal{B}_K$ as the collection of all real numbers $0 < x < 1$ for which the sequences (12) contain only bounded (by a given number K) series with the same binary symbol. If all of the difference sequences $\bar{x}^{(k)}$ belong to $\mathcal{B}_K$, the $\bar{x}$ is called [2] β_K sequence. Some necessary and sufficient conditions a sequence $\bar{x}$ to be a β_K sequence can be found in [2].

The Eggleston theorem ([7]) states that

$$dim\,(\{x \in (0,1) : \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} x_i = p\}) = H(p) \tag{13}$$

where notation $dim(E)$ stands for Hausdorff dimension of set E and

$$H(x) = x \log_2 \frac{1}{x} + (1-x) \log_2 \frac{1}{1-x} \qquad (0 < x < 1) \tag{14}$$

is Shannon entropy function. This function can also be derived from theory of number partitions: if $C(s, N)$ denotes the total number of compositions of number N into s parts,

$$N = m_1 + m_2 + \cdots + m_s \; , \tag{15}$$

then since $C(s, N) = C_{N-1}^{s-1}$, using Sterling formula for binomial coefficients, it can be easily obtained that

$$H(x) = \lim_{N\to\infty} \frac{1}{N} \log_2 C(xN, N) \qquad (0 < x < 1) \; . \tag{16}$$

Analogously, given $K \geq 1$ we define

$$H_K(x) = \lim_{N\to\infty} \frac{1}{N} \log_2 C_K(xN, N) \qquad (0 < x < 1) \tag{17}$$

where $C_K(s, N)$ denotes ([9]; [10], Ch. 4.2) the total number of compositions (15) satisfying the restriction $m_i \leq K$. To be correct in these definitions, one should counts x is rational – if $x = p/q$, then in (16) and (17) N is of the form $N = qn$ and $n \to \infty$. Then the function (16) (as well as the (17)) can be extended on the whole unit interval $0 < x < 1$. Concerning the relation (16) and the procedure just used for defining the Shannon entropy, see also [36, 37].

The next Theorem 2 estimates the Hausdorff dimension of the sets

$$E(p) = \{x \in (0,1) : \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} |x_{i+1} - x_i| = p\} \tag{18}$$

$$E_K(p) = \{x \in \mathcal{B}_K : \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} |x_{i+1} - x_i| = p\} \tag{19}$$

where $0 \leq p \leq 1$ and $K \geq 1$ are arbitrary.

Theorem 2 *The next inequalities*

$$dim\ (E(p)) \leq H(p), \qquad dim\ (E_K(p)) \leq H_K(p) \tag{20}$$

are true.

The following result establishes some relationship between system's γ-characteristic and Hausdorff dimension of its conjugate attractor.

Corollary 1 *If a deterministic system generates the binary sequences (the binary β_K sequences), then for Hausdorff dimension of the attractor $\mathcal{A}$ we have*

$$dim(\mathcal{A}) \leq H(\gamma) \qquad (dim(\mathcal{A}) \leq H_K(\gamma)) \tag{21}$$

where γ is the system's response characteristic defined by Eq. (9).

The H in (21) is Shannon entropy function, having a simple analytic form (14). We do not have such a simple expression for the function H_K. The generating function of numbers $C_K(s, N)$ from (17), by means of which the H_K is defined, is well known ([10], Ch. 4.3):

$$\sum_{N=0}^{\infty} C_K(s, N) q^N = (q + q^2 + \cdots + q^K)^s \ ;$$

using this formula, as in [10], Ch. 4.3, it can be obtained

$$C_K(s, N) = \sum_{p=0}^{s} (-1)^p C_s^p C_{N-K-1}^{s-1} \ .$$

Some asymptotic relations for this quantity can be found in [11] (see [10], Ch. 4, Comments; the work [11] remained unavailable to us). They can be used to derive the explicit analytic form of the function $H_K(x)$. In respect of inequality (20) we note also the work [29], where the positive discrepancy of fractal measures from Shannon entropy is discussed.

2.3 Conjugate orbits and bifurcation points

The conjugate orbits in fact magnify the small fluctuations of the process. The fluctuations are usually occur at unstable branch points [5, 6]. The notion of conjugate orbit allows to determine that each branch point on the bifurcation diagram of the original system can be treated as the source of chaos for some 'shifted' conjugate system.

The computations were made for logistic map $T : x \to rx(1-x)$. Namely, let

$$b_1 < b_2 < \cdots < b_k < \cdots$$

be the bifurcation points of the orbit $X = (x_n)_{n=1}^{\infty}$ of iterates $x_{n+1} = T(x_n)$ $(n \geq 1)$, numbered in increasing order. It is known [15], that

$$\lim_{k\to\infty} b_k = b_\infty = 3.5699\ldots \quad .$$

Given N_k tending to infinity we consider the shifted sequences $\bar{Y}_k = (x_{n+N_k})_{n=1}^{\infty}$. The claim is that if N_k goes to ∞ quickly enough, then the orbits $\bar{\nu}_k$, conjugate to $\bar{Y}_k$, demonstrate the chaotic behavior at points $b_1, b_2, \ldots, b_k$, while they are identically zero for all the different values of control parameter r belonging to interval $(0, b_\infty)$.

This is certain algorithmic formalism of the mentioned in Section 1 descriptive remarks from [5] and [6] relating to small fluctuations. The same analysis can be implemented for many other systems exhibiting the bifurcations. Thus, the bifurcation diagrams of Poincare displacement of Duffing equation ([13], Ch. 11.5), Rossler system ([41], p.46), and forced magnetic oscillator ([12], Ch. 2) are very similar to that of the considered logistic map ([14], Ch. 3).

The bifurcations are often treated as the chaos precursors. Hence, any methods for their detection are of a great practical interest (see, e.g. [12]).

3 Random processes

3.1 Sequences of random independent variables

For the case of random processes the difference analysis has richer consequences and permits theoretical study. We consider discrete random processes of the form

$$\bar{\xi} = (\xi_1, \xi_2, \ldots, \xi_n, \ldots) \tag{22}$$

coordinates ξ_n of which take binary values 0 and 1 with some positive probabilities,

$$P(\xi_n = 0) = p_n, \qquad P(\xi_n = 1) = q_n \qquad (p_n + q_n = 1)\ .$$

Then the differences $\xi_n^{(k)}$,

$$\xi_n^{(k)} = |\xi_{n+1}^{(k-1)} - \xi_n^{(k-1)}| \qquad (n, k \geq 1,\ \bar{\xi}_n^{(0)} = \xi_n)$$

also take binary values with some positive probabilities

$$P(\xi_n^{(k)} = 0) = p_n^{(k)}, \qquad P(\xi_n^{(k)} = 1) = q_n^{(k)} \qquad (p_n^{(k)} + q_n^{(k)} = 1)\ .$$

Hence, one can consider the random difference processes

$$\bar{\xi}^{(k)} = (\xi_1^{(k)}, \xi_2^{(k)}, \ldots, \xi_n^{(k)}, \ldots) \ ; \tag{23}$$

we will also deal with the corresponding random variables of the form

$$\bar{\xi}^{(k)} = 0.\xi_1^{(k)}\xi_2^{(k)} \ldots \xi_n^{(k)} \ldots \qquad (k \geq 0,\ \xi_n^{(0)} \equiv \xi_n,\ \xi^{(0)} \equiv \xi)$$

(the notation is the same as for sequences (23)). It is easy to see that

$$\bar{\xi}^{(k)} = \mathcal{R}^{(k)}\ \bar{\xi}$$

where $\mathcal{R}^{(k)}$ is k-th iterate of the "fractal map" $\mathcal{R}$ from [2]:

$$\mathcal{R}\ \bar{\xi} = 0.\ \xi_1 \oplus \xi_2\ \xi_2 \oplus \xi_3 \ldots \xi_n \oplus \xi_{n+1} \ldots$$

(we use notation $\alpha \oplus \beta$ $(= |\alpha - \beta|)$ for logical sum of binary variables α and β).

We are interested in the limiting behavior of these differences when k goes to infinity. We say that $\bar{\xi}^{(k)}$ converges to a random sequence $\bar{\xi}^{(\infty)}$ if the $p_n^{(k)}$ tend to some numbers $p_n^{(\infty)}$ as $k \to \infty$ and $k \in \Lambda$ (convergence by probability); here Λ is a given infinite subset of natural series and the final probabilities may depend on Λ, $p_n^{(\infty)} = p_n^{(\infty)}(\Lambda)$. Then

$$\bar{\xi}^{\infty} = \bar{\xi}_{\Lambda}^{(\infty)} = (\xi_1^{(\infty)}, \xi_2^{(\infty)}, \ldots, \xi_n^{(\infty)}), \qquad \xi_n^{(\infty)} = \xi_n^{(\infty)}(\Lambda)$$

is again a discrete random process, binary components of which take values 0 and 1 with some final (or stationary – if follow the terminology of Markov processes) probabilities,

$$P(\xi_n^{(\infty)} = 0) = p_n^{(\infty)}, \qquad P(\xi_n^{(\infty)} = 1) = q_n^{(\infty)} \qquad (p_n^{(\infty)} + q_n^{(\infty)} = 1)\ .$$

The following 3 statements are the basic tools we apply for studying the limiting behavior of differences of binary random sequences. For $\epsilon_i \in \{0,1\}$ we use the notation

$$< \epsilon_0, \epsilon_1, \ldots, \epsilon_k >= (\sum_{i=0}^{k} \epsilon_i C_k^i)\ mod\ (2)\ .$$

Lemma 1 *For probabilities of k-th ($k \geq 0$) differences we have*

$$P(\xi_n^{(k)} = \lambda) = \sum_{<\epsilon_0, \epsilon_1 \ldots \epsilon_k> = \lambda} P(\xi_n = \epsilon_0) P(\xi_{n+1} = \epsilon_1) \cdots P(\xi_{n+k} = \epsilon_k) \tag{24}$$

where $n \geq 1$ and $\lambda \in \{0,1\}$ are arbitrary.

It is convenient to represent the probabilities in the form

$$P(\xi_n = \lambda) = \frac{1}{2}(1 + (-1)^{\lambda}\pi_n), \qquad P(\xi_n^{(k)} = \lambda) = \frac{1}{2}(1 - (-1)^{\lambda}\pi_n^{(k)})$$

where $-1 \leq \pi_n, \pi_n^{(k)} \leq 1$ are some numbers. The proof of Lemma 2 is based on the following remark:

Remark 1 *The identity*

$$\sum_{<\epsilon_0,\epsilon_1,\ldots,\epsilon_k>=\lambda} x_0^{\epsilon_0} x_1^{\epsilon_1} \cdots x_k^{\epsilon_k} = \frac{1}{2}[1+(-1)^{\lambda} \prod_{0\le i\le k,\ \alpha_{i,k}=1}^{k} \frac{1-x_i}{1+x_i}] \prod_{i=0}^{k}(1+x_i)$$

is true.

The following lemma gives an explicit expression of $\pi_n^{(k)}$ by means of π_n. To formulate it, we consider the binary analogy $P = (\alpha_{i,k})_{i=0,n;\ \ k=0,\infty}$ of Pascal triangle of binomial coefficients, is defined as follows: $\alpha_{0,k} = \alpha_{k,k} = 1$ and

$$\alpha_{i,k} = \begin{cases} 0, & C_k^i \quad \text{is even} \\ 1, & C_k^i \quad \text{is odd} \end{cases} \qquad \text{i.e.} \quad \alpha_{i,k} = \alpha_{i-1,k-1} \oplus \alpha_{i,k-1}\ .$$

for $1 \le i \le k-1$. The fractal graphical image of the triangle P can be found in [15], where it is considered in connection of cellular automata theory.

Lemma 2 *For arbitrary $n, k \ge 1$ the equality*

$$\pi_n^{(k)} = \prod_{0\le i\le k,\ \alpha_{i,k}=1} \pi_{n+i} \tag{25}$$

is true.

We are interested in the existence of limit of the $\pi_n^{(k)}$ when $k \to \infty$. When k converges to infinity arbitrarily , this limit may not exist. For instance, in the simplest case $p_n \equiv p$ (or $\pi_n \equiv \pi$), it follows from Remark 2 below, that if the binary code of number k contains exactly m units, then $\pi_n^{(k)} = \pi^{2^m}$. Since for every given s the collection Λ_m of all such numbers k is infinite, it is clear that the limit mentioned, generally speaking, does not exist. On the other hand, arbitrary given π_n the Lemma 2 in principle allows to describe all of the infinite subsets $\Lambda \subset N$ for which the limit

$$\pi_n^{(\infty)} = \pi_n^{(\infty)}(\Lambda) = \lim_{k\to\infty,\ k\in\Lambda} \pi_n^{(k)}$$

exists. In other words, Lemma 2 provides a sufficient tool to describe all of the Λ for which the limiting random sequence $\xi^{(\infty)}(\Lambda)$ exists: it follows from (25) that for any infinite Λ and $n \ge 1$ the final probabilities can be computed by formula

$$\ln \frac{1}{|\pi_n^{(\infty)}(\Lambda)|} = \lim_{k\to\infty,\ k\in\Lambda} \sum_{0\le i\le k,\ \alpha_{i,k}=1} \ln \frac{1}{|\pi_{n+i}|} \tag{26}$$

provided the right hand limit exists (or is infinite).

The most of further results relate to studying these final processes for some particular case: we consider the limiting transition of the differences $\bar{\xi}^{(k)}$ as $k \to \infty$ and $k \in \Lambda_0$ where

$$\Lambda_0 = \{k = 2^p - 1 : p = 1, 2, \ldots\}\ .$$

Then for arbitrary k we have all of the $\alpha_{i,k} = 1$ and hence the relation (25) gains a simpler form

$$\pi_n^{(k)} = \prod_{i=0}^{k} \pi_{n+i}\ . \tag{27}$$

From where immediately follows:

Theorem 3 *The limiting process $\bar{\xi}^{(\infty)}(\Lambda_0)$ is the symmetric random walk, i.e. $\pi_n^{(\infty)} \equiv 0$, if and only if when*

$$\sum_{n=1}^{\infty} \ln \frac{1}{|\pi_n|} = \infty \ . \tag{28}$$

In the contrary case, when this series is convergent, we have

$$|\pi_n^{(\infty)}||\pi_1^{(\infty)}|^{-1} = \prod_{i=1}^{n-1} |\pi_i|^{-1} \qquad (= |\pi_n||\pi_{n+1}||\pi_{n+2}|\cdots) \ . \tag{29}$$

It is clear that if $p_n \equiv const$, then (28) holds. If $\pi_n = 2^{-n}$ $(n \geq 1)$, we have an example of a self-conjugate system: $\bar{\xi}^{(\infty)} \equiv \bar{\xi}$. For Poisson distribution, when $p_n = e^{-\lambda}\lambda^n/n!$, using (27) it can be easily computed

$$|p_n^{(\infty)} - p_n| = o\,(p_n) \qquad (n \to \infty) \ .$$

The same relation is valid also for Poisson homogeneous events flow with probabilities $P_n(t) = e^{-\lambda t}(\lambda t)^n/n!$. Indeed, (following [16], Ch. 6.5) for a given t we divide time interval $[0, t]$ into equal intervals of lengths $1/n$ and on the obtained finite lattice consider Bernoulli's trial scheme with success probability is equal to $\lambda t/n$. Since the success probability in s trials tends (as $n \to \infty$) to the function $P_n(t)$, we obtain

$$|P_n^{(\infty)}(t) - P_n(t)| = o\,(P_n(t)) \qquad (n \to \infty) \ . \tag{30}$$

The next statement is an immediate consequence of Theorem 3.

Corollary 2 *The random sequence $\bar{\eta} = (\eta_1, \eta_2, \eta_3, \ldots)$ is a limiting (for some $\bar{\xi}$, i.e. $\bar{\eta} = \bar{\xi}_{\Lambda_0}^{(\infty)}$) process, if and only if when either $\bar{\eta}$ is the symmetric random walk or the sequence $|\pi_n(\bar{\eta})|$ monotone increases to 1 as $n \to \infty$.*

Corollary 3 *If $\bar{\eta}$ is a limiting random sequence (for some $\bar{\xi}$, i.e. $\bar{\eta} = \bar{\xi}_{\Lambda_0}^{(\infty)}$), differing from symmetric random walk, then the random variable*

$$0.\eta_1\eta_2\eta_3\ldots$$

has a pure singular probability distribution.

The latter statement follows from Marsaglia's results [17] – Marsaglia's criterion a random variable $\bar{\eta}$ possesses a singular distribution, is

$$\sum_{n=0}^{\infty} |\pi_n(\bar{\eta})|^2 = \infty \ .$$

The Corollary 2 provides a stronger condition: $|\pi_n(\bar{\eta})| \to 1$. Since we have

$$1 - |\pi_n| = 2\min(p_n, q_n) \ ,$$

the convergence of series in (28) is equivalent to condition

$$\sum_{n=1}^{\infty} \min\,(p_n, q_n) < \infty$$

According to [17], this is also equivalent to a requirement the random variable $\bar{\eta}$ from Corollary 3 has a discrete probability density function.

If $\pi_n \equiv const$, then according to Theorem 3 the $\bar{\xi}^{(\infty)}_{\Lambda_0}$ is the symmetric random walk. It can be shown that for this case Theorem 3 remains valid when the limiting transition is taken over some "wide" subsets Λ of natural series:

Theorem 4 *If $\pi_n \equiv const$, then there exists a set $\Lambda \subset N$ with density 1 in natural series, such that $\bar{\xi}^{(\infty)}_{\Lambda}$ is the symmetric random walk.*

Here, as such a Λ it can be chosen a sequence of natural numbers, for which the total number of units in their binary codes increases to infinity quickly enough. Theorem 4 follows from the next proposition:

Remark 2 *(1) The total number of units in k-th line of binary Pascal triangle is equal to $2^{m(k)}$ where $m(k)$ is the total number of units in binary code of number k. (2) There exist sets $\Lambda \subset N$ such that $\lim_{k\to\infty,\ k\in\Lambda} m(k) = +\infty$ and $dens(\Lambda) = 1$.*

For a given $\bar{\xi}$ we let

$$\gamma(\bar{\xi}, k) = \xi_1^{(k)} + \xi_2^{(k)} + \cdots + \xi_{k-1}^{(k)} \tag{31}$$

– the total number of units in (the realizations of) k-th difference sequence $(\xi_1^{(k)}, \ldots, \xi_{k-1}^{(k)})$; it also coincides with the total number of changes of the binary symbol in the sequence $(\xi_1^{(k-1)}, \ldots, \xi_k^{(k-1)})$:

$$\gamma(\bar{\xi}, k) = |\xi_1^{(k-1)} - \xi_2^{(k-1)}| + \cdots + |\xi_{k-1}^{(k-1)} - \xi_k^{(k-1)}|\ . \tag{32}$$

The next two statements follow from Theorems 3, 4 and Remark 2 and relate to γ-characteristic of random sequences: as in (9), given infinite $\Lambda \subset N$, we define

$$\gamma(\bar{\xi}, \Lambda) = \lim_{k\to\infty,\ k\in\Lambda} \frac{\gamma(\bar{\xi}, k)}{k} \tag{33}$$

where the existence of the limit is assumed. We note again that the limit in (33) with $\Lambda \equiv N$, generally speaking may not exist and then one has to study the corresponding non-trivial cluster sets of the ratio in (33). The same situation holds also for Lyapunov exponent, is defined by Eq. (10).

We consider some particular case of processes $\bar{\xi}$ from (22), requiring

$$P(\xi_n = 0) \geq P(\xi_n = 1) \tag{34}$$

or what is the same, $\pi_n \geq 0$ for all $n \geq 0$.

Theorem 5 *Let us have a random process $\bar{\xi} = (\xi_n)_{n=1}^{\infty}$ satisfying (34). If*

$$\sum_{n=0}^{\infty} P(\xi_n = 1) < \infty \quad \textit{then with probability 1} \quad \lim_{k\to\infty,\ k\in\Lambda_0} \frac{\gamma(\bar{\xi}, k)}{k} = 0$$

and if

$$\sum_{n=0}^{\infty} P(\xi_n = 1) = \infty \quad \textit{then with probability 1} \quad \lim_{k\to\infty,\ k\in\Lambda_0} \frac{\gamma(\bar{\xi}, k)}{k} = \frac{1}{2}\ .$$

Corollary 4 *If $\pi_n \equiv \pi$ then there exists $\Lambda \subset N$ such that $dens(\Lambda) = 1$ and $\gamma(\bar{\xi}, \Lambda) = 1/2$.*

3.2 Conjugate attractors of identically distributed sequences

On considering the random sequences (22), one may assume ([16], Ch. 8.6 and [22], Ch. 4.3) that

$$\xi_n = \xi_n(\omega) = \omega_n \quad \text{where} \quad \omega = 0.\omega_1\omega_2\omega_3\ldots\quad ; \tag{35}$$

here ω is real number from unit interval, given in form of its binary expansion. We are interested in Hausdorff dimension of the sets

$$M = M(p) = \{\omega \in [0,1] : \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \xi_k(\omega) = p\} \tag{36}$$

where is assumed $P(\xi_n = 0) = p = const$. If $p = 1/2$, then M is the set of so-called Borel normal numbers ([7]) for which $mes(M) = 1$ and $dim(M)$ =1 (mes is Lebesgue measure on $[0,1]$). Hence, if in (36) $p \neq 1/2$, then $mes(M) = 0$. However, in that case there exists a singular measure ν on interval $[0,1]$ such that $supp(\nu) = M$, $\nu(M) = 1$ ([7], Ch. 4). The Eggleston theorem provides us with explicit expression for Hausdorff dimension:

$$dim(M(p)) = H(p) \tag{37}$$

where H is Shannon function (14). The derivative of distribution function of measure ν, i.e. the probability density of the random variable

$$0.\xi_1(\omega)\xi_2(\omega)\xi_3(\omega)\ldots$$

is a singular function, concentrated on a set of zero Lebesgue measure. The self-affine graph of such a function can be found in [7], Ch 3.1.

For a given stochastic system it is interesting to describe all of the attractors $\mathcal{A}_m$ (we assume they are numbered by increase of their Hausdorff dimension), mentioned in section 2.1. It can be easily done for the case $\pi_n \equiv const$ $(= \pi)$. Indeed, if we let

$$\Lambda_m = \{k \in N : k = 2^p + m,\ p = 0, 1, 2, \ldots\} \tag{38}$$

then according to Remark 2 for the limiting process $\bar{\xi}^{(\infty)}_{\Lambda_m}$ we have $\pi_n^{(\infty)}(\Lambda_m) \equiv \pi^{2^m}$. For the terms of conjugate orbit $\bar{\nu}$ we have

$$\nu_k = 0.\xi_1^{(k)}\xi_2^{(k)}\xi_3^{(k)}\ldots\ .$$

From where it is clear that the conjugate attractor $\mathcal{A}_m$ corresponding to random variable $\bar{\xi}^{(\infty)} = \bar{\xi}^{(\infty)}_{\Lambda_m}$, coincides with the set, on which the sequence

$$\bar{\xi}^{(\infty)}_{\Lambda_m} = 0.\xi_1^{(\infty)}\xi_2^{(\infty)}\xi_3^{(\infty)}\dots$$

is localized (with probability 1). Then Eggleston theorem implies

Theorem 6 *For the sequence of identically distributed random variables, $\pi_n \equiv \pi$, the set of all its conjugate attractors is a collection of some Eggleston sets,*

$$\mathcal{A}_m = M(\frac{1+\pi^{2^m}}{2}) \qquad (m = 0, 1, 2, \dots)\ . \tag{39}$$

If we let

$$\hat{H}(x) = H(\frac{1+x}{2}) \qquad (\hat{H}(x) = \hat{H}(-x), \quad 0 \le \hat{H}(x) \le 1), \quad -1 \le x \le 1$$

then Theorem 6 gives $dim(\mathcal{A}_m) = \hat{H}(\pi^{2^m})$, hence $\hat{H}^{-1}(dim(\mathcal{A}_m)) = \pi^{2^m}$ and thus the next statement is true:

Theorem 7 *For Hausdorff dimensions of conjugate attractors $\mathcal{A}_m$ of identically distributed random sequence the equalities*

$$(\hat{H}^{-1}(dim(\mathcal{A}_1))^{1/2} = (\hat{H}^{-1}(dim(\mathcal{A}_2))^{1/4} = \dots = (\hat{H}^{-1}(dim(\mathcal{A}_m))^{1/2^m} = \cdots \tag{40}$$

hold.

3.3 Binary Markov processes

Here we give analogies of some results from Sec. 3.1 for the case of infinite binary Markov chains.

Remark 3 *Let $\bar{\xi} = (\xi_n)_{n=1}^{\infty}$ be a binary Markov chain with the transaction probabilities $\pi_n(x,y) = P(\xi_n = y|\xi_{n-1} = x)$ and with the probabilities $p_n(x,y) = P(\xi_n = x)$ of attainment the value x for n steps by the finite chain $(\xi_k)_{k=1}^{n}$. Then for every $k \ge 1$ the difference sequence $\bar{\xi}^{(k)} = (\bar{\xi}_n^{(k)})_{n=1}^{\infty}$ is also a Markov chain and for the corresponding probabilities $\pi_n^{(k)}(x,y)$ and $p_n^{(k)}(x)$ we have the following recurrent relationships:*

$$p_n^{(k)}(x) = p_{n-1}^{(k)}(0)\pi_n^{(k)}(0,x) + p_{n-1}^{(k)}(1)\pi_n^{(k)}(1,x) \tag{41}$$

$$\pi_n^{(k)}(x,y) = \pi_{n-1}^{(k-1)}(0,x)\pi_n^{(k-1)}(x,|x-y|) + \pi_{n-1}^{(k-1)}(1,1-x)\pi_n^{(k-1)}(1-x,1-|x-y|) \tag{42}$$

For the general Markov processes, the computation of distributions $\pi^{(\infty)}$ and $p^{(\infty)}$ for the limiting processes $\bar{\xi}_{\Lambda}^{(\infty)}$ can be complicated. However, we are able to compute the distribution $p^{(\infty)}$ for homogenous processes and for arbitrary given $\Lambda = \Lambda_s = \{2^p + s : p = 0, 1, 2, \dots\}$. This is based on the next two statements.

Lemma 3 *If $\bar{\xi} = (\xi_n)_{n=1}^{\infty}$ is a homogenous Markov process and $\Lambda = \{2^p : p \geq 0\}$ then for the difference processes $\bar{\xi}^{(k)} = (\xi_n^{(k)})_{n=1}^{\infty}$, $k \in \Lambda$, we have*

$$P(\xi_m^{(k)} = \lambda) = q^k \sum_{\epsilon,\delta \in \{0,1\}} ((\frac{q}{1-q})^{\epsilon+\delta-1} \sum_{[p/2]=\lambda} \sum_{k=1}^{\infty} C_{m-p}^{k-\epsilon-\delta} C_{p-1}^{k-1} x^p y^k \tag{43}$$

where $\lambda \in \{0,1\}$, $x = \frac{s}{q}$, $y = \frac{1-s}{s}\frac{1-q}{q}$, $s = p(1,1)$, $q = p(0,0)$ and $p(x,y)$ is the transaction probability function for the process $\bar{\xi}$.

Remark 4 *The next identity*

$$\sum_{p=0}^{\infty} \sum_{k=0}^{\infty} C_p^k C_{m-p}^k x^p y^k = (1+x)^m T_m(\frac{x(y-1)}{(1+x)^2}) \tag{44}$$

where $T_m(z) = \sum_{k=0}^{m} C_{m-k}^k z^k$ is Chebyshev polynomial, is true.

These three statements imply:

Theorem 8 *If $\bar{\xi} = (\xi_n)_{n=1}^{\infty}$ is a homogenous Markov process and $\Lambda = \{2^p : p \geq 0\}$, then we have*

$$\lim_{k \in \Lambda,\ k \to \infty} P(\xi_n^{(k)} = \lambda) = |\lambda - \frac{2(1-s)(1-q)}{(1-s)+(1-q)}| \tag{45}$$

where λ, s, and q are the same as in Lemma 3.

It is clear from (42) that if $\bar{\xi}$ is homogenous process then for every $k \geq 1$ the difference process $\bar{\xi}^k$ is again homogenous. It is clear now that the limits of the quantities $P(\xi_n^{(k)} = \lambda)$ can be computed for all the sets of indices of the form $\Lambda_s = 2^p + s : p = 0, 1, 2 \ldots\}$. Indeed, it is not difficult to see from Theorem 8, that this limit is

$$\lim_{k \in \Lambda_s,\ k \to \infty} P(\xi_n^{(k)} = \lambda) = |\lambda - \frac{2\pi_0^{(s)}(1,0)\pi_0^{(s)}(0,1)}{\pi_0^{(s)}(1,0) + \pi_0^{(s)}(0,1)}| \tag{46}$$

It is also clear that the quantities $\pi_0^{(s)}$ are easily determined recurrently according to (41) and (42) (and hence their explicit computation is also possible). Thus we have the convergence of the quantities $P(\xi_n^{(k)} = \lambda)$ along the Λ_s, however we cannot confirm the convergence for the probabilities $\pi_n^{(k)}(x,y)$. If their convergence is known apriori, the relation (41) implies

$$\frac{\pi_n^{(\infty)}(0,1)}{\pi_n^{(\infty)}(1,0)} = |1 - \frac{1}{2}\frac{\pi_n^{(0)}(0,1) + \pi_n^{(0)}(1,0)}{\pi_n^{(0)}(1,0)\pi_n^{(0)}(0,1)}| \tag{47}$$

– this ratio can measure the 'strength' of Markov dependence property for the final process.

3.4 One-dimensional diffusion

On obtaining the relations (30) for Poisson events flow, we used the circumstance that this process can be approximated by some discrete random processes. By the same way, one can obtain such a result for the diffusion processes. However, in this case we need to consider some "averaged" differences $\hat{\xi}_n^{(k)}$, are defined as

$$P(\hat{\xi}_n^{(k)} = 0) = \frac{1}{2}(1 - \hat{\pi}_n^{(k)}) \quad \text{where} \quad \hat{\pi}_n^{(k)} = (\pi_n^{(k)})^{1/k} = (\pi_n \pi_{n+1} \dots \pi_{n+k})^{1/k} \ .$$

We will also need to use the $(C,1)$ summation by Cesaro [18]:

$$(C,1) \int_x^{+\infty} f ds = \lim_{z \to +\infty} \frac{1}{z} \int_x^z f ds \ .$$

Indeed, let the diffusion process $\bar{\xi}(t)$ is determined by Fokker-Plank-Kolmogorov equation ([16, 19, 20])

$$\frac{\partial f(x,t)}{\partial t} = -b \frac{\partial f(x,t)}{\partial x} + \frac{a}{2} \frac{\partial^2 f(x,t)}{\partial x^2} \tag{48}$$

where $a = a(x)$ and $b = b(x)$ are some coefficients and $f(x,t)$ is the probability density of diffusing particle. It is well known (see e.g., [16], Ch. 5.4 and [19]) that such processes can be obtained as a result of some limiting transition from random walk on one-dimensional lattice: we consider the lattice L_h with a step $h > 0$ and assume that a particle changes its position on the discrete time moments are proportional to some small $\tau > 0$ and with the probabilities

$$p = \frac{1}{2}(1 - \pi), \qquad q = \frac{1}{2}(1 + \pi) \quad \text{where} \quad \pi = \frac{1}{2} \frac{b}{a} h \ . \tag{49}$$

Imposing the restriction $h^2/\tau \to a$ (it can be assumed $a \equiv 1$) and based on Central Limit Theorem, one can deduce the FPK-equation as well as to obtain its solution.

The scheme we apply to study the difference structure of such continuous processes is the following. If we have a discrete motion $\bar{\xi}_h$ on the lattice L_h, the transition probabilities of which permit an equality of the form

$$\pi_n = C_n h \tag{50}$$

then, after calculations by formula (27), we transform the quantities $\pi_n^{(k)}$ for probabilities of k-th difference process taken from random walk $\bar{\xi}_h$ to a form

$$\hat{\pi}_n^{(k)} = C_{n,k} h \quad (= C_{x,k} h)$$

where the coefficients $C_{x,k}$ are such that there exists the $\lim_{k \to \infty} C_{x,k} = C_x$. In such a way we can consider a random walk, corresponding to the discretized difference process, assigned on the same lattice L_h and with the same time scale τ. For the considering case (49) we have $\hat{\pi}_n^{(k)} = \pi_n = bh/2$, i.e. $\hat{\xi}_n^{(k)} \equiv \bar{\xi}_h$. From where we conclude that $\hat{\xi}^{(\infty)}(t)$ coincides with $\bar{\xi}(t)$. One can see, the same arguments lead to the following formula:

$$|B(x)| = \exp(- \ (C,1) \int_x^{\infty} \ln |b(z)| dz) \ . \tag{51}$$

By such a way, the next statement is true:

Theorem 9 *The limiting averaged differential process $\hat{\xi}^{(\infty)}(t)$, taken from the diffusion $\bar{\xi}(t)$ is again a diffusion process. If $\bar{\xi}$ satisfies equation (48) then the drift coefficient B for $\hat{\xi}^{\infty}$ can be computed by (51).*

The same analysis is also applicable to general birth and death processes [16] as well as to abstract diffusion processes, considering in non-standard stochastic analysis [30] including the Ising model. We note another possible applications of the above given approach. It is queuing analysis in computer networks (see e.g., [20, 21]). This theory deals with randomly arriving demands to some processors. The random process of time intervals between arrival moments is studied. On investigating of so-called heavy traffic [20, 21], a great importance has the diffusion approximation – the process of queue lengths can be approximated by diffusion process (48) (cp. Wiener's neuron from next section). This area of applications is as large, that requires an independent study.

4 Neural networks

4.1 Stochastic neuron models

Wiener's model of neuron ([8], p.299) has been suggested by Mandelbrot and Gerstein (see [23, 40]). On the studying the actual neuron spike trains, they found that for some instances these trains are well approximated by (bounded) diffusion process. The idea is that the excitatory and inhibitory signals, arriving to neuron's input, can be mathematically interpreted as a (bounded) random walk on one-dimensional lattice with a small constant step h. The limiting (as $h \to 0$) process $\bar{\xi}$ is described by Ito's stochastic differential equation ([19], Ch. 5.4)

$$d\bar{\xi}(t) = \mu(x,t)dt + \sigma(x,t)d\bar{W}(t) \tag{52}$$

where $\bar{W}$ is Wiener's process. These authors have found that for some appropriately chosen values of parameters μ and σ, the process $\bar{\xi}(t)$ fits well the experimental data on neural activity. The equation (52) is equivalent to diffusion equation (48) with $\mu \equiv b$ and $\sigma^2 \equiv a$ ([8], p.299).

It is noted ([8], p.299) that inclusion a noisy component to a given deterministic system may cause the diffusion process in the system. When the noise is of a small intensity, the system shows the Poisson behavior. By this reason, in some cases the neural activity can be described by Poisson driven ([23, 40] and [8], p.299) stochastic models.

The next proposition is the main statement of this subsection. It is simply a reformulation of some results (Theorem 9 and Eq. (30)) from Section 3.

Corollary 5 *The system, conjugate (in "average" sense) to Wiener neuron is again a Wiener neuron. The system, conjugate to Poisson driven neuron is again an (asymptotically) Poisson driven neuron.*

Applying the approach described in Sec. 3.4, the analogous result might also be stated for Ornstein-Uhlenbeck neuron ([8], p.299), that is defined as a diffusion process with time dependent coefficients a and b (or μ and σ) of some special form.

4.2 New type of memory in neural networks

We consider the update equation ([8], pp.119, 230, 930), that describes the dynamics of neural network consisting of n McCulloch-Pitts neurons

$$x_k(t+1) = \sigma(h_k(t) - \theta_k) \quad \text{where} \quad h_k(t) = \sum_{j \neq k} w_{k,j} x_j(t) \tag{53}$$

Here $w_{k,j}$ are synaptic strengths, θ_k are threshold constants, σ is activation function, h_k is synaptic potential, variable x_k stands for k-th neuron binary states, and t designates the discrete time. Different choices of σ, so-called sigmoid functions, are possible. It is accepted to include the probabilistic "noisy" term to this equation ([8], p.930),

$$P(x_k(t+1) = \delta) = \frac{1}{2}(1 + (-1)^{\delta} \pi_k) \ ; \tag{54}$$

here, it can be chosen, e.g. ([8], p.902)

$$\pi_k = tanh[T^{-1} h_k(t)]$$

where the variable $T > 0$ (temperature) reflects the level of noise.

The main two problems, investigating in neural networks are retrieval and learning problems. The first one studies the dynamics of neural states x_k provided the connections $w_{i,j}$ are time independent and fixed. The most interest consists in revealing the attractors of dynamics (54), which are considered as the memory storage of given network. The basic result is the Hopfield theorem, establishing that under some restrictions on matrix W the configuration point

$$\bar{x} = \bar{x}(t) = (x_1(t),\ x_2(t), \ldots, x_n(t), \ldots) \tag{55}$$

converges to some fixed-point attractors. The Hopfield nets consist of spin neurons, taking values ± 1. It is proposed [28] the existence of some 'energy' function ([8], p.363) associated with W

$$E = E(W) = -\frac{1}{2} \sum_{i \neq j} w_{i,j} s_i s_j \tag{56}$$

which, provided certain restrictions, permits the Lyapunov function ([8], pp.363, 230): the value of E is increased with any update of spins. The local minima points of the function E are treated as the attractor memory states. This memory can take about 20% of the whole configuration space [28]. Some other works in neural networks ([8], p.258), generalizing the Hopfield's approach, introduce and study the networks of chaotic elements. The aim is to provide the existence of an hierarchical memory storage of coexisting attractors.

The analysis from previous sections reveals a new type of attractors of the dynamics (54) and therefore, a new type of memory in neural networks. Indeed, for arbitrary k-th neuron we consider its states

$$\bar{\xi}_k = (x_k(1),\ x_k(2), \ldots, x_k(t), \ldots)$$

defined by (55), as a discrete random process. Despite of we were mostly dealt with independent random variables, the results of Section 3.3 show that the difference analysis

is also applicable to the case of binary random Markov processes - an important restriction (see e.g., [8, 27, 28]), usually imposing on the process (53) (or (54)) of the brain states. The relation (26) in principle allows to compute all of the final processes $\bar{\xi}_k^{(\infty)}(\Lambda)$, corresponding to those $\Lambda \subset N$, for which the final probabilities $\pi^{(\infty)}(\Lambda)$ exist. The discrete processes

$$(\bar{\xi}_1^{(\infty)}(\Lambda_1),\ \bar{\xi}_2^{(\infty)}(\Lambda_2), \dots \bar{\xi}_n^{(\infty)}(\Lambda_n), \dots) \tag{57}$$

can be treated as some final (or stationary) processes of the network dynamics, given by Eqs. (54) and (55). The Cartesian products

$$\mathcal{A}_{s_1}^{(1)} \times \mathcal{A}_{s_2}^{(2)} \times \cdots \times \mathcal{A}_{s_k}^{(k)} \times \cdots \tag{58}$$

$(1 \leq s_i \leq \infty)$ where $\mathcal{A}_n^{(k)}$ is an attractor for the conjugate orbit, on which the random variable $\bar{\xi}_k^{(\infty)}(\Lambda_n)$ is localized, are the attractors for configuration point (55) and therefore can be treated as a new type of memory of a given neural network.

The learning or task-adapted problems, considering in neural networks theory, have an inverse statement: given set of patterns of configuration points to determine the matrix W for which the fixed point attractors coincide with this set of patterns. It is supposed that some finite number of given patterns to be learned by network, as well as their components, are random and have identically distributed components ([8], p.651). In the framework of the analysis presented, the learning problem can be stated as follows: given random processes of the form

$$\bar{\eta} = (\eta_1, \eta_2, \dots, \eta_n, \dots) \tag{59}$$

to determine the matrix W of adjustable connections in such a way that these $\bar{\eta}$ coincide with some final processes (58) with some $\Lambda_i \subset N$.

4.3 Elimination of noise

The real numbers are sometimes represented in the "pulse density system" ([31]) – e.g., the limit in the relation (36) represents the number p in this system. The density of finite number of alternating signals is easily expressed through the densities of its constituents. In this system, all the basic operations with signals are also possible ([31]). This representation is also used on the studying the multidimensional stochastic systems (e.g., on investigating the baker's transformation [33]). The difference attractor $\mathcal{A}$ for such complex system reflects the synthetic information on all the constituents. In respect of information alteration, due to the timing of the signals, see also [8], p.693.

We have (see Eqs. (31) and (33))

$$\gamma(\bar{\xi}, \Lambda) = \lim_{k \to \infty, k \in \Lambda} \frac{1}{k} \sum_{i=0}^{k-1} \xi_i^{(k)}$$

and therefore, the γ is a type of density. It is mentioned in [31] a device (the charge capacitor), transforming the pulse density to some analog quantity. This remark in fact indicates a way to set a correspondence between the γ-coefficient of neural activity and

neuron's electrical characteristics. Hence, the γ can be treated as some mixed analog-digital characteristic of neural activity – according to von Neumann [31], the neuron enables to combine the analog and digital features in its activity.

We use these remarks to deduce a new type of neuron firing condition, based on analysis from Sec. 3 and formulated in probabilistic terms. Accepted condition for neuron firing says that the total electrical charge in neuron should exceed some threshold level. In the formal neural networks, it is reflected in the update equation (53) – a weighted sum of input signals should exceeds some threshold level. On the other hand, the work [31] claims that the actual firing conditions may have a very different form. The next suggestion is derived from consideration of discrete random processes. One can see, it has the theoretical-probabilistic character and does not refer to any neural context.

Let us have a neuron, receiving on its input the signals from other neurons. As in the previous section, we assume that these signals are some independent random variables η_k, taking the values $+1$ (excitatory signal) and -1 (inhibitory signal). We assume also that we deal with discrete-time process and that the η_k are ordered by the growth of their arrivals time. Letting $\xi_k = (1+\eta_k)/2$, we have on the input of neuron a random binary process $\bar{\xi}$. Such processes have been considered in Sec. 3 and therefore all the results of this section can be applied. We are interested in the statement of Theorem 5, which we now reformulate as follows:

Corollary 6 *Let $\bar{\xi} = (\xi_n)_{n=1}^{\infty}$ be a random binary sequence with independent terms ξ_n, satisfying (34). If*

$$\sum_{n=0}^{\infty} P(\xi_n = 1) < \infty \quad \textit{then with probability 1} \quad H(\gamma\,(\bar{\xi})) = 0\ ,$$

and if

$$\sum_{n=0}^{\infty} P(\xi_n = 1) = \infty \quad \textit{then with probability 1} \quad H(\gamma\,(\bar{\xi})) = 1\ .$$

Let us now have infinite number of some binary random variables (neurons) $\xi_k(t)$ with given distribution of probabilities

$$P(\xi_k(t) = 0) = p_k(t), \quad P(\xi_k(t) = 1) = q_k(t) \qquad (p_k(t) + q_k(t) = 1)\ . \tag{60}$$

Here $k \geq 1$ is neuron's number and $t = 0, 1, 2, \ldots$ is the discrete time. If we understood the realization of the random variable $\xi_k(t)$ as

$$\xi_k(t) = \begin{cases} 1 & \text{k-th neuron is active (fired) at moment t} \\ 0 & \text{k-th neuron is inactive (silent) at moment t}\ , \end{cases}$$

it can be said that every probabilistic neural net is a random process $\mathcal{N}_P$ of the form

$$\bar{\xi}(t) = (\xi_1(t), \xi_2(t), \ldots, \xi_n(t), \ldots) \tag{61}$$

Now we introduce a deterministic network $\mathcal{N}_D$, associated with the process $\mathcal{N}_P$. Given 2 infinite matrices: the stochastic matrix $P = (p_k(t))_{k,t=0}^{\infty}$ of the probabilities (60), and the binary matrix of the connections $W = (w_{i,j})_{i,j=0}^{\infty}$:

$$w_{i,j} = \begin{cases} 1 & \text{i-th neuron affects on j-th neuron} \\ 0 & \text{i-th neuron does not affect on j-th neuron} \end{cases}$$

we assign the evolution equation of the net $\mathcal{N}_D$ as follows: It is clear that the process

$$\bar{\xi}(k;t) = (w_{k,1}\xi_1(t), w_{k,2}\xi_2(t), \ldots, w_{k,n}\xi_n(t), \ldots)$$

consisting of all the neurons, affecting on k-th at the moment t, is the input process for k-th neuron. According to Corollary 6 the quantity

$$\Gamma(\bar{\xi}(k;t)) = H(\gamma(\bar{\xi}(k;t)))$$

($\Gamma = Ho\gamma$ is the composition of H and γ) is either 0 or 1. We impose the following firing condition for the neurons of $\mathcal{N}_D$:

$$x_k(t+1) = \Gamma(\bar{\xi}(k;t)) \ . \tag{62}$$

Let us explain these definitions. The equation (62) is the evolution equation of the process $\bar{x}(t)$: given $\bar{\xi}(t)$ it allows to compute the $\bar{x}(t+1)$. This means we have required the γ-characteristic (and hence the entropy $H(\gamma)$) to be the basic quantity, determining the dynamics of the deterministic network $\mathcal{N}_D$: in order to determine its state at next time step (i.e. $\bar{x}(t+1)$), each neuron of $\mathcal{N}_P$ computes the γ-characteristic and then the entropy $H(\gamma)$ of its present input (the $\bar{\xi}(k;t)$).

It follows from Corollary 6, that the dynamics of the net $\mathcal{N}_D$ can be represented in the form

$$x_k(t+1) = \Delta(S_k(t)) \quad \text{where} \quad S_k(t) = \sum_{j\neq k} w_{k,j}x_j(t)p_j(t) \tag{63}$$

(cp. Eq. (53)) where Δ is the impulse function of the form

$$\Delta(x) = \begin{cases} 1 & x = \infty \\ 0 & x \neq \infty \ . \end{cases}$$

In other words, we have proved the following

Theorem 10 *For every neural network $\mathcal{N}_P$, consisting of probabilistic neurons $\xi_k = \xi_k(t)$ there exists some deterministic neural network $\mathcal{N}_D$ which at each time t computes the γ-characteristic of the inputs of ξ_k.*

The probabilistic "noisy" networks has been introduced by W. Little [27] in order to include to the theoretical studies the noise actually presenting in the brain. The Theorem 10 shows, that when we are interested in the entropy aspects of neural activity and if we follow the difference approach, the noise can be eliminated. In this respect, the latter theorem can also be treated as some analogy of de Leeuw-Moore-Shannon-Shapiro statements [39] from the automata theory for the case of neural networks. Note also that Theorem 10 makes the modified neural networks remarkably closer to actual brain: the brain enables to operate avoiding the influence of external noise.

References

[1] A. Yu. Shahverdian and A. V. Apkarian, *Fractals*, 7, 1, 1999.

[2] A. Yu. Shahverdian, *Fractals*, 8, 1, 2000.

[3] A. Yu. Shahverdian and A. V. Apkarian, Periodic response of periodically perturbed stochastic systems, LANL preprint, 2000

[4] A. J. Lichtenberg and M. A. Lieberman, *Regular and Stochastic Motion*, Springer, N-Y, 1983.

[5] I. Prigogine *From Being to Becoming*, Freeman, San Francisco, 1980.

[6] Yu. L. Klimontovich, The Turbulent Motion and Structure of Chaos, Nauka, Moscow, 1990 (in Russian).

[7] P. Billingsley *The Ergodic Theory and Information*, Wiley, New York, 1965.

[8] *The Handbook of Brain Theory and Neural Networks*, ed. M. A. Arbib, The MIT Press, Massachusetts, 1995.

[9] J. Riordan *Combinatorial Identities*, Wiley, N-Y, 1968

[10] G. E. Andrews *The Theory of Partitions*, Addison-Wesley, Massachusetts, 1976

[11] Z. Star *Aequations Math*, 1976

[12] F. C. Moon, *Chaotic Vibrations*, Wiley, 1987.

[13] V. Barger and M. Olsson, *Classical Mechanics: Modern Perspective*, 2-nd ed., McGraw-Hill, 1995.

[14] H. Schuster, *Deterministic Chaos*, Springer, 1984.

[15] U. Quasthoft *in: Fractals in Physics. Proc. VI Trieste Int. Symp. on Fractals in Physics, ICTP, Trieste, Italy, 1985. ed. L. Pietronero and E. Tosatti. 1986*

[16] W. Feller *An Introduction to Probability Theory and its Applications*, N-Y, 1950.

[17] G. Marsaglia *Ann. Math. Stat.* 42, 6, 1971.

[18] G. H. Hardy *Divergent Series*, Oxford, 1949

[19] Yu. V. Prokhorov and Yu. A. Rozanov *Theory of Probabilities*, Moscow, Nauka, 1987

[20] G. P. Basharin, P. P. Bocharov, Ya. A. Kogan *Queuing Analysis for Computer Networks. Theory and Computational Methods.* Moscow, Nauka, 1989.

[21] E. G. Jr. Coffman and M. I. Reiman *Diffusion approximations for computer and communications systems* in: Math Computer Performance and Reliability, eds. G. Iazeolla, P. J. Courtois, A. Hordijk, Amsterdam: North-Holland, 1984, p.33-53.

[22] A. N. Shiryaev *The Probability*, Moscow, Nauka, 1980

[23] G. L. Gerstein and B. Mandelbrot *Random walk models for the spike activity of a single neuron*, Biophys. J., 4: 41-68, 1964

[24] K. Weisenfeld and F. Moss, *Nature*, 373, 33-36, 1995.

[25] F. M. Moss, A. Bulsara, M. F. Shlesinger *J. Stat. Phys., 1/2, Proc. NATO Advanced Research Workshop on Stochastic Resonance in Physics and Biology*, 1993.

[26] D. R. Chialvo and A. V. Apkarian *J. Stat. Phys.*, 70, 375-391, 1993.

[27] W. A. Little *Math. Biosciences*, 19, 1974, 101-120

[28] J. Hopfield *Proc. Natl. Ac. Sci.* 79:2554-2558, 1982

[29] A. Beghdadi, C. Andraud et al. *Entropic and multifractal analysis of disordered morphologies*, Fractals, 1, 3, 1993.

[30] S. Albeverio, J. E. Fenstad, R. Hoegh-Krohn, T. Lindstrom *Nonstandard Methods in Stochastic Analysis and Mathematical Physics*, Acad. Press, London, 1986.

[31] J. von Neumann *The Computer and The Brain*, New Haven, Yale Univ. Press, 1958.

[32] C. E. Shannon *Communication in the presence of noise*, Proc. IEEE, v.72, N 9, 1984

[33] N.F.G. Martin and J.W. England *Mathematical Theory of Entropy*, Cambridge Univ. Press, 1984

[34] A. Yu. Shahverdian LANL preprint, 1999

[35] A. Yu. Shahverdian, Russ. Math. Surveys, v. 47, 1992

[36] F. Reike, D. Warland, R. de Ruyter van Steveninck, W. Bialek *Spikes. Exploring the Neural Code*, The MIT Press, Massachusetts, 1997

[37] N. Chomsky and G. A. Miller *Finitary models of language users*, in: Handbook of Mathematical Psychology, vol. 2, Wiley, N-Y, 1963.

[38] Symposium. The design of machines to simulate the behavior of the human brain, Trans. IRE, EC-5, N 4, 1956.

[39] *Automata Studies*, ed. C. E. Shannon and J. McCarthy, Princeton Univ. Press, 1956.

[40] L. Glass and M. Mackey *From Clocks to Chaos. The Rhythms of Life* Princeton Univ. Press, 1988

[41] G. Gouesbet et al. *Ann. N-Y Ac. Sci.*, v.808, 1997.